



Dixon, P., Dakin, H., & Wordsworth, S. (2016). Generic and disease-specific estimates of quality of life in macular degeneration: Mapping the MacDQoL onto the EQ-5D-3L. *Quality of Life Research*, 25(4), 935-945. <https://doi.org/10.1007/s11136-015-1145-x>

Peer reviewed version

License (if available):
Unspecified

Link to published version (if available):
[10.1007/s11136-015-1145-x](https://doi.org/10.1007/s11136-015-1145-x)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Springer at <http://link.springer.com/article/10.1007%2Fs11136-015-1145-x>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Generic and disease-specific estimates of quality of life in macular degeneration: Mapping the MacDQoL onto the EQ-5D-3L

Padraig Dixon^{1*}, Helen Dakin², Sarah Wordsworth²,

1 School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, United Kingdom

2 Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Old Road Campus, Oxford, OX3 7LF, United Kingdom

* Corresponding author

Email of corresponding author: padraig.dixon@bristol.ac.uk

Telephone number of corresponding author: 00 44 117 928 7319

Keywords: MacDQoL, mapping algorithm, EuroQoL EQ-5D, age-related macular degeneration.

1 **Abstract**

2 **Purpose:** The Macular Degeneration Quality of Life (MacDQoL) instrument is a
3 validated condition-specific measure of quality of life in patients with macular
4 degeneration. This paper presents the first mapping algorithm to predict EQ-5D from
5 responses to the MacDQoL instrument.

6 **Methods:** Responses to the MacDQoL and EQ-5D-3L instruments from 482 patients
7 were collected from the IVAN multicentre trial of two alternative drug treatments for
8 neovascular age-related macular degeneration. Regression specifications were
9 estimated using OLS, censored least absolute deviation, Tobit and two-part models.
10 Their predictive performance was assessed using mean squared error. An internal
11 validation sample based on a random selection of 25% of patients was used to
12 assess the performance of the model estimated on the remaining 75% of patients.

13 **Results:** A two-part model had the best predictive performance on the full sample.
14 The covariates of this model include responses and weighted impact scores for all 23
15 condition specific domains of the MacDQoL, and responses to a general MacDQoL
16 quality of life question. The selected models were successful at predicting means and
17 standard deviations of target populations, but prediction is weaker at the upper and
18 lower extremes of the EQ-5D-3L distribution.

19 **Conclusion:** The mapping algorithms provide a means of predicting EQ-5D-3L index
20 scores from MacDQoL scores, and could facilitate cost-effectiveness analyses when
21 the latter but not the former are available to researchers. Further validation of the
22 performance of the algorithms using external data would provide a means of
23 establishing the robustness of the algorithms.

24 Introduction

25 There is growing interest from health economists and quality-of-life researchers in
26 understanding the relationship between condition-specific and generic measures of
27 patient outcome, especially where the former but not the latter are collected within
28 clinical trials and other study contexts [1-3]. Condition-specific measures of health-
29 related quality of life (HRQoL) may offer greater sensitivity than generic measures [4].
30 However, for health care commissioners, generic measures are useful in providing
31 comparable information on effectiveness and cost-effectiveness across different
32 interventions and patient groups.

33 There is evidence that validity, sensitivity and responsiveness of generic measures
34 differs between diseases [5]. Recent studies have suggested that the responsiveness
35 of generic HRQoL measures in visual disorders may be limited [2; 5] although the
36 evidence base in this area is small. Our paper examines the relationship between
37 generic and condition-specific measures of HRQoL in macular degeneration. Age-
38 related macular degeneration is a progressive chronic eye disease. It is a leading
39 cause of irreversible blindness and visual disability, accounting for approximately
40 50% of all vision impairments or blind registrations in developed countries [6].

41 Prevalence is expected to increase significantly in the developing world with
42 demographic change leading to aging populations [7]. In the UK, 39,700 new cases
43 of neovascular macular degeneration are estimated to occur each year, with an
44 estimated prevalence of 263,000 cases nationally [8]. Estimated prevalence at
45 different age groupings in the UK ranges from 0.1% of men and women aged 50-54
46 to 24.7% of men aged over 90 and 25.6% of women [8]. Given the large number of
47 patients with the condition, it is important to explore the HRQoL improvements

48 associated with various treatments for the condition and assess how incremental
49 cost-effectiveness compares with interventions for other conditions in order to assess
50 where and how limited healthcare resources should be best deployed.

51 The MacDQoL (**Mac**ular **D**egeneration **Q**uality of **L**ife) [9] is a macular degeneration-
52 specific outcome measure. The MacDQoL is considered acceptable to patients and
53 is related to several vision measures [10]. Its test-retest reliability has been confirmed
54 [11], as have its metric properties in UK, French, German, Italian and American
55 populations [12]. Translations or adaptations are available in 15 languages [9].
56 There is currently no mechanism to estimate EQ-5D values from responses to the
57 MacDQoL instrument. This is a potential limitation of the instrument as health state
58 preference values are required to inform cost-effectiveness analyses in technology
59 appraisals and drug reimbursement decisions. While direct elicitation studies require
60 recruitment of a new sample of patients or citizens answering cognitively challenging
61 questions, mapping studies can be conducted on existing datasets. The objective of
62 our paper was to use trial data on patients with age-related macular degeneration to
63 produce a set of coefficients that can be used to reliably transform patients'
64 responses on the MacDQoL questionnaire to EQ-5D-3L index scores. It is
65 understood to the first example of MacDQoL being mapped onto a generic utility
66 instrument. This paper presents mapping algorithms that predict generic EQ-5D-3L
67 index scores from the MacDQoL.

68 **Methods**

69 The data were derived from the IVAN trial [13; 14], a UK factorial randomised
70 controlled trial which assessed ranibizumab versus bevacizumab for the treatment of
71 macular degeneration. Participants aged 50 and above had active but previously
72 untreated neovascular age-related macular degeneration in the study eye, and were
73 randomised to receive ranibizumab or bevacizumab, and either discontinuous
74 treatment or continuous monthly injections for two years. IVAN participants
75 completed both MacDQoL and EQ-5D-3L at three time points.

76 **HRQoL instruments**

77 The MacDQoL consists of two general overview questions and 23 specific questions
78 or domains. The two general items (DQOL-I and DQOL-II) are scored individually and
79 measure generic present quality of life (DQOL-I), and macular degeneration-specific
80 quality of life (DQOL-II) [10]. Each of the 23 specific domains contains an 'impact'
81 question¹ and an 'importance' question². The use of an importance-weighted impact
82 allows for an overall weighted impact of each item to be calculated as the product of
83 an impact rating and an importance rating, with a range from -9 (maximum negative
84 impact of macular degeneration) to +3 (maximum positive impact of macular
85 degeneration). The average weighted impact score is calculated from the sum of
86 weighted ratings specific domains divided by the number of applicable domains, and
87 also has a range from -9 to +3. The use of both individualised impact and importance
88 responses reflects how MacDQoL was designed to capture the quality of life impacts

¹ Impact response categories are answered in response to questions such as 'If I did not have macular degeneration, friendships and social life would be...'. Typical categories are 'very much better' [scoring -3], 'much better' [scoring -2], 'a little better' [scoring -1], 'the same' [scoring 0], and 'worse' [scoring +1].

² Importance response categories are answered in response to questions such as 'My friendships and social life are..' with response categories 'very important' [scoring 3], 'important' [scoring 2], 'somewhat important' [scoring 1], 'not at all important' [scoring 0].

89 of macular degeneration, rather than simply measuring only the functional impact of
90 the condition of visual performance. A more extended discussion of the instrument is
91 presented in [10] and in the online supplementary material.

92 EQ-5D-3L is a standardised, generic measure of health status that provides a
93 descriptive profile of health states, and a single index value for health status [15]. It is
94 the preferred utility instrument of the National Institute for Health and Care
95 Excellence (NICE) [16]. The index value is anchored on a maximum value of 1 for
96 perfect health, 0 to represent death, and negative values that reflect states worse
97 than death. It contains 5 dimensions: mobility; self-care; usual activities; pain and
98 discomfort; and anxiety and depression. The three level version (EQ-5D-3L)
99 measures these dimensions by three categories, corresponding to no problems,
100 some problems, and extreme problems.

101 **Data** The 610 adults participating in IVAN self-completed large print EQ-5D-3L
102 questionnaires in UK English (with assistance from research nurses where
103 necessary) at baseline, at 3, 12 and 24 months, at study exit and after serious
104 adverse events. EQ-5D-3L responses were valued using the UK time trade-off tariff
105 [17].

106 MacDQoL was administered in English by telephone at 3, 12 and 24 months during
107 the IVAN trial. Data from all time points were eligible for inclusion in the analysis.
108 Four domains of the MacDQoL instrument allowed people to respond 'Not applicable'
109 (or similar), instead of answering the impact/importance questions: work, personal
110 relationships, family life, and holidays. These values were recoded to zero (to
111 indicate no impact on quality of life) so as to be retained for analysis in for some of
112 the estimated models; this is described in more detail in the online appendix

113 (Supplementary Material 1). Observations with missing data on any other MacDQoL
114 or any of the five EQ-5D-3L index questions were excluded from the analysis.

115 The MacDQoL instrument was due to be administered 14 days after EQ-5D-3L to
116 reduce patient burden but on 1,008 occasions (88% of all instances for which paired
117 data on the date of EQ-5D-3L and MacDQoL are available), it was completed more
118 than 14 days after the EQ-5D, with a mean difference of 41 days later than protocol.
119 Only observations where the MacDQoL was completed no more than 90 days after
120 EQ-5D-3L completion, and no more than 1 day before EQ-5D-3L was due to be
121 administered, were used in our mapping analysis in order to ensure consistency
122 between the EQ-5D-3L and the MacDQoL instruments. Sensitivity analyses varying
123 this cut-off, and further details of this issue, are reported in the online appendix
124 (Supplementary Material 1). After excluding patients who did not complete both the
125 EQ-5D-3L and MacDQoL within the specified timeframe, and the 20 observations
126 with missing data on one or more EQ-5D-3L or MacDQoL questions, the regression
127 analyses were estimated with up to 817 observations from 462 patients (70% of
128 those randomised) out of 858 MacDQoL responses (95% of all completed MacDQoL
129 questionnaires included in the analysis).

130 **Statistical methods**

131 The objective of the analysis was to estimate mean EQ-5D-3L values from participant
132 responses to the MacDQoL instrument. To assess the degree of overlap between the
133 two instruments, we first estimated Spearman's rho (rank correlation coefficients)
134 showing the correlations between EQ-5D-3L and DQOL-I, DQOL-II and the
135 MacDQoL using the `-spearman-` Stata command.

136 Regression models using different specifications and different estimators were
137 applied to the data in order to identify models that predicted EQ-5D-3L successfully,

138 as determined by mean squared error (MSE) and mean absolute error (MAE). There
139 is no clear guidance as to which of MSE and MAE is preferable [18]. MSE was used
140 as the primary measure here because it penalises deviations from the mean more
141 heavily than does the MAE. Four estimators that have been used in previous
142 mapping studies [3] were applied and their predictive performance compared:
143 ordinary least squares (OLS); Tobit; centred least absolute deviations (CLAD); and
144 two-part models.

145 The starting point for this analysis was the estimation of linear OLS models. Linear
146 models are widely used in mapping studies [3], and offer a useful basis for
147 comparison with more complex models. They also retain the possibility of being
148 predictively successful in their own right. OLS predictions of EQ-5D-3L values >1
149 were recoded to 1. A criticism of OLS models is that they do not reflect the natural
150 'ceiling' of EQ-5D-3L values at 1, and require this ex-post recoding adjustment.
151 Alternative estimators that account directly for the upper limit on EQ-5D-3L were
152 therefore also assessed.

153 Tobit models [19] allow for continuous dependent variables that have constrained
154 ranges, and therefore can allow for the censoring of EQ-5D-3L at 1. The model
155 estimates, using maximum likelihood, a continuous distribution for values of EQ-5D-
156 3L index scores below the ceiling value of 1, while placing a positive (discrete)
157 probability value on the censored outcome of EQ-5D=1. The rationale for applying
158 Tobit is to allow the estimator to reflect the natural ceiling of EQ-5D-3L at perfect
159 health (when EQ-5D-3L=1) rather than to model the censoring to which Tobit models
160 are usually applied [19]. Pullenayegum et al [20] note that Tobit (and CLAD – see
161 below) models may be biased, when the intention is to perform economic evaluation,
162 if it is assumed that 'true' utility extends beyond 1. However, any potential bias of an
163 estimator is of secondary concern in the context of this study, where we are aiming to

164 accurately predict mean EQ-5D-3L index scores, rather than to produce, for example,
165 unbiased estimates of covariates that are believed to be causally associated with
166 quality of life.

167 We also estimated the CLAD models [21] using the `-clad-` command in Stata [22].
168 CLAD models, which minimise absolute deviations from the median, are similar to
169 Tobit models in their capacity to account for the upper limit in EQ-5D-3L as a
170 dependent variable in regressions, but differ from Tobit models in being consistent
171 even if error terms are not normal and having standard errors that are robust to
172 heteroskedasticity.

173 The final set of estimators comprised two-part models. These models account for the
174 upper limit of EQ-5D-3L by separately modelling individuals at the upper limit of EQ-
175 5D, and modelling those individuals below this limit. Unlike Tobit, CLAD and OLS
176 models, the two-part model uses two different estimators to produce a single
177 estimate. Specifically, two-part models estimate in the binary first ‘part’ a logistic
178 regression to predict the probability of having an EQ-5D-3L score equal to exactly 1
179 ($\text{Probability(Utility} = 1)$), and in the second part use an OLS model to predict utility (U)
180 for those individuals with less than perfect health (i.e. EQ-5D-3L scores <1). The
181 second part of the two-part model therefore regresses the EQ-5D score on the same
182 or a closely related set of covariates as used in the first part on those patients with
183 less than perfect health. For the OLS part of the two-part model model, predictions of
184 EQ-5D-3L of greater than 1 were again recoded to 1. Predictions are produced by
185 combining the predictions of the first-part logit model with those of the second-part
186 OLS model by weighting the OLS prediction by the probability of having imperfect
187 health using the following formula:

$$188 \quad \text{Predicted Utility} = \text{Probability(Utility}=1) + (1 - \text{Probability(Utility}=1)) * U$$

189 Seemingly unrelated regression (SUR), implemented with the `—suest—` command in
190 Stata, was used to adjust the standard errors of the two-part model to allow for
191 correlation between the error terms of the two parts of the model.

192 For each estimator, we explored the performance of a number of different model
193 specifications with using different scores or responses from MacDQoL. We explored
194 some models in which each MacDQoL domain was coded using one ordinal variable
195 measuring response level. We also evaluated whether weighted impact scores (the
196 product of impact and importance) improved EQ-5D-3L predictions. However, the
197 latter two specifications treat the MacDQoL responses as though they were
198 continuous, implicitly assuming that the levels between response levels are evenly
199 spaced. We therefore also explored the impact of using dummy variables for different
200 levels of the ordinal response variables, which recognises that the variable has a
201 categorical nature. Further specifications considered the effect on predictive
202 accuracy of recoding or excluding questions with an ‘N/A’ response.

203 MacDQoL impact variables had limited numbers of responses in some of the most
204 extreme categories, which meant that econometric models including dummy
205 variables for all response levels dropped variables. To address this issue, responses
206 were recoded for models using dummy variables so that the most extreme responses
207 were collapsed into adjacent categories: the ‘Very much better’ and ‘Much better’
208 categories were merged, as were the ‘Worse’ and ‘The same categories’. Only 14
209 MacDQoL responses (1.6%) indicated that macular degeneration *improved* quality of
210 life in any domain. Estimating models without merging these categories therefore
211 tended to produce collinearity and could not be estimated (see online supplementary
212 material). It may be worth investigating this issue in the future if a much larger
213 dataset becomes available for analysis. The recoded domain responses used in the
214 analysis therefore had three levels.

215 We excluded from consideration model specifications that did not have convergent
216 likelihood, which excluded large parts of the sample (because of low levels of data on
217 one variable) and/or which dropped variables because of perfect collinearity.
218 Coefficients with $p > 0.05$ and/or with counterintuitive signs were retained in model
219 specifications given that the primary objective of the estimated models was to
220 maximise prediction accuracy.

221 Twenty two different model specifications were evaluated for each of the four
222 different estimators, producing up to 88 MSE values. A list of combinations of model
223 specification and estimator is presented in Tables 4 and 5 of the online appendix.
224 The appropriateness of model specification was determined by reference to
225 predictive accuracy. The predictive accuracy of the chosen specification were also
226 assessed by examining Q-Q plots of actual versus predicted EQ-5D index scores,
227 scatter plots of actual versus predicted scores, and simple distributional comparisons
228 (see online supplementary material).

229 All models were estimated using Stata version 13.1 (Statacorp: College Station,
230 Texas) using clustering to adjust standard errors to allow for repeated observations
231 made on individual patients.

232 **Validation**

233 Ideally, the prediction accuracy of a mapping algorithm should be tested on an
234 external dataset to assess whether the model is 'over-fitted' to noise in the data
235 rather than the underlying relationship, and to evaluate how the model will perform in
236 other datasets. External data were not available in this case, so an internal validation
237 sample was used [1]. This was conducted by using the (pseudo)-random number
238 generating function in Stata to select a 25% 'validation' sample ($n=115$) leaving an

239 'estimation' sample with the remaining 75% of data (n=367). This percentage split is
240 typical of that used in other studies [23].

241 Results

242 Summary statistics

243 Participants in the IVAN trial had a mean age of 77.7 years, of whom 60% were
 244 female. Participants needed a best corrected distance of visual acuity of at least 25
 245 letters on the Early Treatment Diabetic Retinopathy Study chart to be eligible for
 246 inclusion, and the mean at baseline was 61.4 letters [13]. Some 14% of individuals
 247 randomised had angina, 19% dyspnoea and less than 10% had experienced one or
 248 both of myocardial infarction or stroke; 64% were current or past smokers [13]. Table
 249 1 presents descriptive statistics for the key variables used in the analysis for this
 250 paper for the 482 patients with no missing data on EQ-5D-3L and MacDQoL domains
 251 used in regression analysis. Further information on the demographic characteristics
 252 of participants in the IVAN trial is available in [13].

253 **Table 1 Summary statistics of HRQoL instruments**

254

	Full sample (N=860 observations from 482 patients)	Estimation sample (N=655 observations from 367 patients)	Validation sample (N=205 observations from 115 patients)
EQ-5D: mean (SD)^a	0.8417 (0.1960)	0.8459 (0.1912)	0.8282 (0.2108)
	Min: -0.181 Max: 1	Min: -0.181 Max: 1	Min: -0.077 Max: 1
MacDQoL weighted impact: mean (SD)	-1.8801 (1.7902)	-1.9670 (1.8352)	-1.6005 (1.61003)
	Min: -9 Max: 0.1363	Min: -9 Max: 0.1363	Min: -7.318 Max: 0
DQOL-I: mean (SD)	1.3934 (0.9947)	1.4167 (1.0115)	1.3186 (0.9371)
	Min: -3 Max: 3	Min: -3 Max: 3	Min: -1 Max: 3
DQOL-II: mean (SD)	-1.4639 (1.0393)	-1.4984 (1.0113)	-1.3529 (1.1198)
	Min: -3 Max: 1	Min: -3 Max: 1	Min: -3 Max: 1

255 a – SD: Standard deviation.

256

257 Some 21% of responses to the DQOL-I question (macular degeneration-specific
 258 quality of life) indicated no negative impact of macular degeneration on HRQoL. EQ-
 259 5D-3L and DQOL-I, DQOL-II and the MacDQoL average weighted impact measure
 260 were highly correlated under Spearman's rho, and the null of independence was

261 strongly rejected in all cases ($p < 0.01$). Lower EQ-5D-3L scores were correlated with
262 lower MacDQoL estimates of quality of life. There was a small cluster of observations
263 with low EQ-5D-3L scores: 0.58% of observations were below 0, and 3.95% were
264 below 0.5.

265 **Mapping MacDQoL onto EQ-5D-3L**

266 The best-performing model differed depending on whether predictive accuracy was
267 assessed using the estimation sample or the validation sample (Supplementary
268 material 1, Tables 4 and 5).

269 The model having the lowest MSE in the estimation sample (Model 4 – see
270 Supplementary material for numbering) was a two-part model that included dummy
271 variables for all 23 MacDQoL domains and continuous interactions of impact and
272 importance for all of these domains in the first (logit) part of the model, while the
273 second (OLS) part of the model included the same set of covariates, plus the generic
274 quality of life measure DQOL-I (treated as a continuous variable).

275 The model with the lowest MSE in the validation sample (Model 5) was an OLS
276 model estimated on all 23 MacDQoL domains treated as dummy variables, and
277 continuous interactions of impact and importance for all of these domains. This is the
278 same set of covariates as the first part of the two-part model selected in the
279 estimation sample.

280 When these models were estimated on the full sample of 817 observations and
281 predictions generated, the two-part Model 4 had a better predictive performance than
282 the OLS Model 5. Generally, it is more appropriate to select models based on their
283 performance in the validation sample to avoid over-fitting and typically the model that
284 performs best in the validation sample also performs best on the full sample.
285 However, in this case, results differ markedly because the validation sample has

286 lower mean EQ-5D-3L scores and prediction errors are especially bad for such
 287 patients. In the validation sample, 9.28% of observations had EQ-5D-3L <0.6,
 288 compared with 6.42% in the estimation sample. Differences in the distribution of EQ-
 289 5D-3L between the two samples may have affected the choice of model and
 290 estimator.

291 **Table 2 Summary comparison of observed and predicted EQ-5D-3L on full, estimation**
 292 **and validation samples**

Variable	N ^a	Mean	Standard deviation	Minimum	Maximum	MSE ^d	MAE ^d
Observed data							
Observed EQ-5D-3L – full regression sample	817	0.8441	0.1937	-0.181	1		
Observed EQ-5D-3L – estimation sample	623	0.8488	0.1886	-0.181	1		
Observed EQ-5D-3L – validation sample	194	0.8290	0.2091	-0.077	1		
Predictions – two-part Model 4^b							
Predicted EQ-5D-3L from two-part model	817	0.8481	0.0844	0.4037	1	0.0293	0.1269
Predicted EQ-5D-3L from two-part model – estimation sample	623	0.8523	0.0944	0.2089	1	0.0256	0.1178
Predicted EQ-5D-3L from two-part – validation sample	194	0.8668	0.0924	0.4511	1	0.0500	0.1631
Predictions - OLS Model 5^c							
Predicted EQ-5D-3L from OLS model – full sample	817	0.8436	0.0785	0.4837	1	0.0310	0.1329
Predicted EQ-5D-3L from OLS model – estimation sample	623	0.8480	0.0880	0.3317	1	0.0273	0.1245
Predicted EQ-5D-3L from OLS model – validation sample	194	0.8619	0.0884	0.4235	1	0.0494	0.1642

293 a – Number of observations: Of the 482 patients included in any regression analysis, 20 were excluded from the
 294 OLS and two-part models shown here because of missing data in EQ-5D-3L scores and MacDQoL domains
 295 variables used in regressions. This reduced the number of available observations by 43 from 860 to 817 for the
 296 largest regression samples.
 297 b - Specification of first part of Model 4: A logistic regression to predict EQ-5D=1 where covariates comprised all
 298 23 MacDQoL domain responses as dummy variables, and all 23 weighted impact variable values as continuous
 299 variables. Specification of second part of Model 4: An OLS regression of EQ-5D-3L values for those patients with
 300 EQ-5D index scores<1, using the same explanatory variables the first part, plus the DQOL-I variable as a
 301 continuous variable.
 302 c - Specification of Model 5: An OLS regression of EQ-5D-3L values in which explanatory variables comprised all
 303 23 MacDQoL domains treated as dummy variables, and continuous interactions of impact and importance for all
 304 of these domains.
 305 d - Values shown in bold are the lowest in their respective samples of those estimated and displayed in the table.
 306

307 We therefore took the conservative approach of presenting both models: i.e. those
308 informed by predictive accuracy of models in both the estimation sample (Model 4)
309 and the validation sample (Model 5). Variance-covariance matrices are available as
310 online appendix (Supplementary Material 2) and can be used to estimate standard
311 errors around predicted values when the mapping algorithm is applied to other
312 samples.

313

Table 3 Coefficients to predict EQ-5D-3L from MacDQoL estimated using the full sample (n=817)

Variable		Two-part Model 4		OLS Model 5
		Logit coefficients (SE)	OLS coefficients (SE)	Coefficients (SE)
Household tasks^a	A little better	-0.0035 (0.3125)	-0.0353 (0.0281)	-0.0200 (0.0239)
	Much better or very much better	0.2789 (0.4947)	-0.0223 (0.0419)	-0.0092 (0.0366)
Personal affairs^a	A little better	-0.7794 (0.3088)	0.0148 (0.0285)	-0.0333 (0.026)
	Much better or very much better	-1.0097 (0.5231)	0.0469 (0.0427)	-0.0107 (0.0413)
Shopping^a	A little better	0.4939 (0.2855)	-0.0113 (0.0253)	0.0256 (0.023)
	Much better or very much better	0.2746 (0.4576)	-0.0544 (0.0383)	-0.0022 (0.0337)
Work^a	A little better	-0.2048 (0.7607)	0.0168 (0.0844)	-0.0003 (0.0583)
	Much better or very much better	-1.5117 (1.3659)	0.0189 (0.0948)	-0.0610 (0.08)
Relationships^a	A little better	0.2794 (0.5866)	-0.0462 (0.0664)	-0.0283 (0.0495)
	Much better or very much better	0.6153 (1.0143)	-0.0321 (0.1257)	0.0128 (0.0902)
Family life^a	A little better	-0.5902 (0.4357)	-0.0655 (0.0374)	-0.063 (0.0323)
	Much better or very much better	-1.0378 (0.7979)	-0.1384 (0.0726)	-0.1320 (0.0685)
Friendships^a	A little better	-0.137 (0.346)	-0.0014 (0.0279)	-0.0078 (0.0265)
	Much better or very much better	0.6822 (0.6614)	-0.044 (0.0495)	0.0105 (0.0481)
Physical appearance^a	A little better	-0.4769 (0.4034)	0.0245 (0.0328)	-0.0045 (0.0291)
	Much better or very much better	-1.0209 (0.8434)	0.0102 (0.0678)	-0.0381 (0.0566)
Physical activity^a	A little more	-0.2113 (0.3348)	0.0543 (0.0316)	0.0159 (0.0274)
	Much more or very much more	-0.9505 (0.555)	0.0316 (0.0541)	-0.0291 (0.0488)
Out and about^a	A little better	-0.7453 (0.3237)	-0.0325 (0.0313)	-0.0567 (0.0282)
	Much better or very much better	-1.5539 (0.5259)	0.0177 (0.0463)	-0.0819 (0.0484)
Holidays^a	A little better	0.1084 (0.2988)	0.015 (0.0281)	0.0186 (0.0231)
	Much better or very much better	-1.035 (0.4759)	0 (0.0401)	-0.0198 (0.0339)
Leisure and hobbies^a	A little more	0.0764 (0.2922)	-0.021 (0.0295)	-0.0029 (0.025)
	Much more or very much more	0.1379 (0.4243)	-0.0296 (0.0402)	-0.0176 (0.0347)
Self-confidence^a	A little better	0.5343 (0.3112)	-0.0033 (0.0271)	0.0171 (0.0235)
	Much better or very much better	1.2402 (0.5675)	-0.0053 (0.0477)	0.0538 (0.0404)
Motivation^a	A little better	0.1446 (0.3171)	0.0535 (0.0263)	0.0363 (0.0239)
	Much better or very much better	0.4464 (0.5721)	0.0581 (0.049)	0.0514 (0.0438)
	A little better	-0.1992 (0.4257)	0.0740 (0.0389)	0.0454 (0.0366)

Reaction of others^a	Much better or very much better	0.32 (0.7181)	0.0354 (0.0776)	0.061 (0.073)
Feelings about the future^a	A little better	0.1088 (0.2805)	-0.0314 (0.0246)	-0.0098 (0.0235)
	Much better or very much better	0.0033 (0.3993)	0.004 (0.038)	0.0063 (0.0323)
Financial situation^a	A little better	0.0984 (0.7161)	-0.0346 (0.0441)	-0.0164 (0.047)
	Much better or very much better	-0.207 (1.6711)	-0.1145 (0.0996)	-0.1115 (0.1009)
Independence^a	A little better	0.3108 (0.3318)	0.0263 (0.0281)	0.0275 (0.0274)
	Much better or very much better	0.3307 (0.5383)	0.0746 (0.0523)	0.0517 (0.0461)
Doing things for others^a	A little better	-0.2579 (0.2987)	0.0457 (0.029)	0.0024 (0.0254)
	Much better or very much better	-0.2808 (0.5)	-0.0079 (0.0495)	-0.0312 (0.0419)
Experiencing mishaps or losing things^a	A little better	0.234 (0.3211)	-0.0607 (0.0299)	-0.0171 (0.0277)
	Much better or very much better	0.9399 (0.6205)	-0.0512 (0.0684)	0.0115 (0.0594)
Enjoyment of meals^a	A little more	-0.2301 (0.3928)	0.0181 (0.0254)	0.0067 (0.0292)
	Much more or very much more	1.0081 (0.647)	0.0185 (0.0606)	0.0656 (0.0576)
Time taken to do things^a	A little less	-0.3971 (0.244)	0.0033 (0.0298)	-0.0264 (0.0223)
	Much less or very much less	-0.1084 (0.4568)	0.0401 (0.0419)	0.0091 (0.0359)
Enjoyment of nature^a	A little more	-0.0565 (0.3128)	0.0027 (0.0251)	0.0001 (0.0231)
	Much more or very much more	-1.0278 (0.4949)	0.0114 (0.0313)	-0.038 (0.0358)
Impact/importance interaction	Household tasks	-0.0378 (0.0816)	-0.0102 (0.0071)	-0.0102 (0.0063)
	Personal affairs	-0.1396 (0.0753)	0.0024 (0.0066)	-0.0022 (0.0062)
	Shopping	0.056 (0.0679)	-0.0049 (0.0058)	0.0016 (0.005)
	Work	-0.3265 (0.228)	0.0084 (0.0221)	-0.0116 (0.0161)
	Relationships	-0.0583 (0.1677)	-0.0129 (0.0204)	-0.0106 (0.014)
	Family life	-0.1342 (0.1198)	-0.0196 (0.0122)	-0.0146 (0.0103)
	Friendships	0.0432 (0.0971)	-0.0116 (0.008)	-0.0028 (0.0076)
	Physical appearance	-0.0846 (0.1306)	0.0022 (0.0107)	-0.0015 (0.0089)
	Physical activities	-0.0975 (0.0946)	-0.0006 (0.0096)	-0.0045 (0.0086)
	Out and about	-0.2568 (0.0768)	0.0074 (0.008)	-0.0105 (0.0079)
	Holiday	-0.1742 (0.0903)	-0.0038 (0.0088)	-0.0098 (0.007)
	Leisure	0.0469 (0.0628)	-0.002 (0.0055)	-0.0006 (0.0048)
	Self-confidence	0.2509 (0.0887)	-0.0016 (0.0075)	0.0107 (0.0067)
	Motivation	-0.0764 (0.088)	0.0153 (0.0079)	0.0043 (0.0073)
	Reaction of others	0.2273 (0.1279)	0.0221 (0.017)	0.0312 (0.0159)
	Feelings about the future	-0.0212 (0.0582)	-0.0032 (0.0058)	-0.0036 (0.0046)
	Financial situation	0.1248 (0.2599)	-0.0151 (0.015)	-0.0084 (0.0158)
	Independence	0.1857 (0.0734)	0.0111 (0.0073)	0.0162 (0.0068)

	Do for others	-0.1099 (0.0792)	-0.0013 (0.0075)	-0.0072 (0.0066)
	Mishaps	0.0835 (0.0957)	-0.0055 (0.0116)	0.0008 (0.0101)
	Enjoyment of meals	0.1699 (0.0946)	0.0106 (0.0119)	0.0163 (0.011)
	Time taken to do things	0.0215 (0.0775)	0.0028 (0.0072)	0.0022 (0.0065)
	Enjoyment of nature	-0.0596 (0.0746)	0.0036 (0.0052)	-0.0005 (0.0055)
DQOL-I			0.0358 (0.0105)	
Constant		0.2991 (0.1705)	0.6701 (0.0264)	0.8781 (0.0146)
Pseudo R² ^b		0.1363		
Adjusted R² ^b			0.0450	0.0940
MSE (full sample)			0.0293	0.0310
MAE (full sample)			0.1269	0.1329

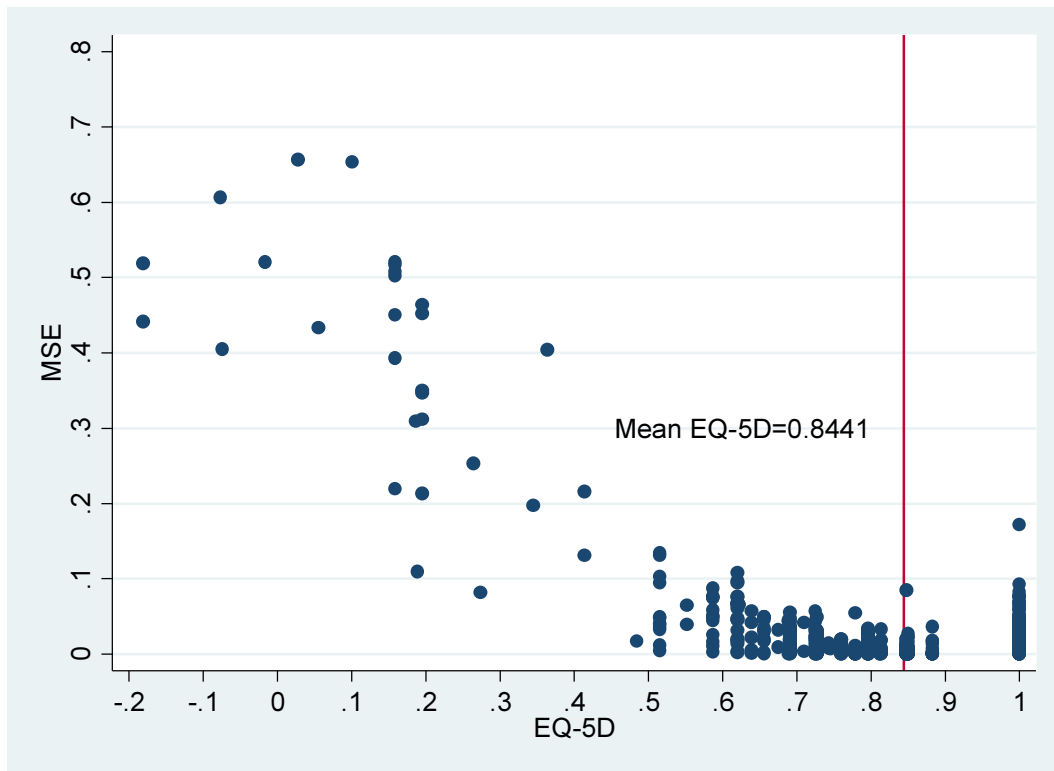
316 a: All 23 items in the MacDQoL instrument begin with 'If I did not have macular degeneration...' The respondent is
317 then asked to consider the impact of the condition on the activities described above and rate it on a five-point
318 scale (worse, same, a little better, much better, or very much better). The two extreme categories have been
319 combined with the second most extreme. The reference response to these questions is 'Same or worse' –
320 meaning that the variables should be interpreted against a patient response that the presence of macular
321 degeneration has no impact on that specific domain or their actions in that domain would be worse without
322 macular degeneration.
323 b: R² values are those associated with the respective regressions used to generate these coefficients – pseudo R²
324 in the case of the logit regression and adjusted R² in the case of the OLS regressions.
325 For the OLS model, EQ-5D-3L values can be calculated from the coefficients in this table by taking the constant
326 term (0.8781) and adding the coefficients for the relevant item responses (e.g. adding -0.02 if the patient would
327 find household tasks a little better if they did not have macular degeneration), then, for each domain, adding on
328 the product of the coefficient for the impact/importance interaction (e.g. -0.0378 for household tasks), that patient's
329 impact score on that domain (e.g. 2 if they would find household tasks a little better) and their importance score
330 (e.g. 3 if household tasks are very important to them). For both OLS and two-part models, the impact scores used
331 to calculate weighted impact should be Worse=0, The same=1, A little better=2, Much better=3, Very much
332 better=4, and for Importance: 0=Not at all important, 1= Somewhat important, 2=Important, 3=Very important.
333 EQ-5D-3L values can be calculated in a similar way for the two-part model. The procedure for the OLS
334 component of the two-part model (U) is as described for the OLS model but using coefficients from the second
335 column of data in the table above. The log odds of having perfect health (utility = 1) can be calculated by taking
336 the constant term (0.2991) and adding on the coefficients for the relevant impact item responses and the product
337 of the coefficient for weighted impact scores, that patient's impact score and their importance score. Log-odds
338 should then be converted to probabilities by taking the exponent of the log odds to calculate the odds of a utility of
339 1 and dividing the odds by 1+odds to give Probability(Utility=1). Predicted utility can then be calculated as:
340 Predicted Utility = Probability(Utility=1) + (1 – Probability(Utility=1))*U
341

342 The purpose of the regressions is to predict EQ-5D-3L index scores rather than
343 estimate the magnitude and sign of coefficients, but the following example shows
344 how these coefficients may be interpreted. The 'Personal affairs' question asks the
345 respondent to state 'If I did not have [macular degeneration], I could handle my
346 personal affairs...'. Using the recoded variables deployed in the analysis, the
347 coefficient of the OLS regression on the 'Better or very much better' dummy of the
348 'Personal Affairs' question is -0.0107, which can be interpreted as suggesting that

349 patients who could manage their personal affairs much better or very much better
350 would have approximately 1% lower EQ-5D-3L scores than patients for whom
351 macular degeneration had no impact on management of their personal affairs. The
352 coefficients cannot be given a causal interpretation, and in any event there is no clear
353 pattern amongst the coefficients that indicate which domains of MacDQoL might
354 have a particular influence on EQ-5D index scores. In the OLS models, coefficients
355 including 'family life' impacts have p-values <0.05 , and likewise the 'personal affairs'
356 impact domain in the logit model. Several coefficients (e.g. shopping on the two-part
357 logit model) have counterintuitive signs, suggesting that EQ-5D-3L scores would be
358 improved if macular degeneration had a larger impact on that domain. However, all
359 variables were kept in the model to avoid reducing prediction accuracy by omitting
360 parts of the source instrument.

361 Prediction errors were high at the extremes of the EQ-5D-3L distribution, particularly
362 the lower end, while errors were smaller near mean EQ-5D-3L: MSE is <0.1 for
363 values within ± 0.15 of mean EQ-5D-3L (Figure 1). This suggests that the accuracy of
364 prediction, as measured by MSE, is smallest when predicting mean EQ-5D-3L
365 compared to other quantiles of EQ-5D-3L. This suggests that both algorithms are
366 better suited to prediction of mean EQ-5D-3L from MacDQoL than for values far from
367 mean EQ-5D-3L index scores.

368 **Fig. 1** MSE at different levels of EQ-5D-3L – full sample, two-part Model 4



369

370 Discussion

371 This paper has presented the methods and results of a mapping algorithm that
372 generates EQ-5D-3L values from responses to the condition-specific MacDQoL
373 instrument. The mapping algorithm can be used to estimate EQ-5D-3L index values
374 in circumstances where only MacDQoL data is available – this circumstance may
375 arise in study designs where the latter but not the former has been collected by
376 researchers. Although MacDQoL is not widely used at present, this may change with
377 increasing research on eye disease. Estimated models are successful at predicting
378 the mean EQ-5D-3L for this specific sample, but may be less successful for
379 predicting individuals' utility (particularly for those with low or very high utility). We
380 found that the predictive performance of the studied models is weaker for patients
381 with EQ-5D-3L values <0.6 than for values ≥ 0.6 . A similar trend has been
382 encountered in other literature concerned with eye disease and HRQoL [24]. For
383 cost-effectiveness analysis, accurate prediction of sample means may be more
384 important than the scores of individuals at particular parts of the utility distribution
385 [25].

386 As this is the first study mapping between these instruments, it was not possible to
387 compare the predictive performance against other published studies, and it is
388 important to test the algorithm presented here on other datasets when they become
389 available. A review of mapping studies in other areas [26] found MAE estimates to
390 range between 0.0011 to 0.19, which encompasses the MAEs found in this study for
391 the selected models (0.1269 for the two-part model and 0.1329 for the OLS model).
392 The only other study mapping predicting EQ-5D-3L for patients with macular
393 degeneration identified in a recent systematic review [3] reported a root mean

394 square error of 0.2163 for its recommended model when mapping from the National
395 Eye Institute Vision Function Questionnaire [27]. The square of this term is 0.0468,
396 which is somewhat larger than the MSEs for preferred models in the different
397 samples (Models 4 and 5) of 0.0293 and 0.0310, although comparisons are affected
398 by the different samples and instruments involved.

399 Our best performing models included some coefficients that were non-significant
400 and/or had counterintuitive signs. This has also been observed in previous studies,
401 including those with much larger samples [28]. Such counterintuitive signs could
402 cause the mapping algorithm to predict a reduction in EQ-5D when the patient's
403 MacDQoL profile has improved and could reflect collinearity within our sample or
404 genuine opposition between domains. We suggest that it is appropriate to include the
405 whole of the source instrument for completeness, and because omitting inconsistent
406 or non-significant coefficients can reduce prediction accuracy [28]. Selecting models
407 based on information criteria may have given a more parsimonious model, but are
408 not appropriate for selecting mapping models, where prediction accuracy is the main
409 concern [1; 29; 30].

410 Participants in IVAN had mean baseline EQ-5D-3L index scores of 0.85, which is
411 higher than estimates of UK age-adjusted mean population norms [31], potentially
412 because of the trial inclusion criteria and the requirement for monthly visits [32]. This
413 is a limitation of the study if the relationship between the two instruments varies with
414 patients' general health. If slope coefficients were to change when estimated on a
415 population in worse health, this would affect the level of QALYs estimated, but not
416 necessarily affect estimates of incremental QALYs or incremental cost-effectiveness.
417 Further validation of the mapping models on an external dataset is required to assess
418 prediction accuracy in other populations.

419 The choice of estimator differed depending on whether model selection was based
420 on MSE in the validation sample or MSE in the estimation sample. This could be
421 explained by the greater proportion of individuals with low EQ-5D-3L scores in the
422 validation sample. An external validation dataset would provide a more robust test of
423 the model's overall predictive accuracy, and may offer evidence regarding the
424 applicability of the algorithm to other types of patient population.

425 The sample size of this study is a limitation. A larger sample size would improve the
426 precision of estimates, and may address the differences between validation and
427 estimation samples as well as the small number of instances of counter-intuitive
428 signs of particular variables in the preferred model specification. Larger sample sizes
429 may also support the application of other potentially relevant approaches such as
430 response mapping [33]. The number of observations available used in this analysis
431 (817) is below the median (1,167) and mean (6,069) number of observations in a
432 recent review [3] of mapping studies.

433 MacDQoL, like other 'DQoL' instruments, asks patients to rate the importance of
434 each domain, as well as rating the impact of their condition. We found that the
435 product of importance and impact affected EQ-5D utility over and above the effect of
436 impact scores. An unexplored issue is whether it is reasonable to map from a
437 disease-specific instrument that uses these patient-reported measures of importance
438 to a generic instrument (EQ-5D) that uses valuations of preferences between health
439 states provided by members of the general public. We note that NICE guidance for
440 technology appraisal [16] recommends the use of a generic instrument valued using
441 public (not patient) preferences, but does not specify the characteristics that ought to
442 be possessed by disease-specific instruments from which EQ-5D index values are
443 obtained via mapping algorithms.

444 **Conclusion**

445 Our paper presents the first set of algorithms to map from MacDQoL to health state
446 values, which will facilitate cost-effectiveness studies in this area. The models are
447 reasonably simple, in that any dataset with complete responses to MacDQoL can be
448 used to predict EQ-5D, with no additional data required. The methods described are
449 also likely to be applicable to similar analyses in other disease areas. The models
450 presented in this paper can be used to estimate mean EQ-5D-3L values in other
451 samples. However, our models have been evaluated only on patients with
452 neovascular age-related macular degeneration with MacDQoL average weighted
453 impact scores between -9 and 0.14, with a median value of -1.36, and future work is
454 required to assess whether our models would perform as well in other patient groups.

- 477 1. Longworth, L., & Rowen, D. (2013). Mapping to Obtain EQ-5D Utility Values
478 for Use in NICE Health Technology Assessments. *Value in Health*, 16(1), 202-
479 210.
- 480 2. Tosh, J., Brazier, J., Evans, P., & Longworth, L. (2012). A review of generic
481 preference-based measures of health-related quality of life in visual disorders.
482 *Value in Health*, 15(100883818), 118-127.
- 483 3. Dakin, H. (2013). Review of studies mapping from quality of life or clinical
484 measures to EQ-5D: an online database. *Health and Quality of Life Outcomes*,
485 11, 6.
- 486 4. Brazier, J., & Dixon, S. (1995). The use of condition specific outcome
487 measures in economic appraisal. *Health Economics*, 4(4), 255-264.
- 488 5. Longworth, L., Yang, Y., Young, T., Mulhern, B., Hernández Alava, M.,
489 Mukuria, C., Rowen, D., Tosh, J., Tsuchiya, A., Evans, P., Devianee
490 Keetharuth, A., & Brazier, J. (2014). Use of generic and condition-specific
491 measures of health-related quality of life in NICE decision-making: a
492 systematic review, statistical modelling and survey. *Health Technol Assess*,
493 18(9).
- 494 6. Usha, C., Jennifer, E., & Philip, J. R. (2010). Age related macular
495 degeneration. *BMJ*, 340.
- 496 7. Woo, J. H., Sanjay, S., & Au Eong, K.-G. (2009). The epidemiology of age-
497 related macular degeneration in the Indian subcontinent. *Acta*
498 *Ophthalmologica*, 87(3), 262-269.
- 499 8. Owen, C. G., Jarrar, Z., Wormald, R., Cook, D. G., Fletcher, A. E., &
500 Rudnicka, A. R. (2012). The estimated prevalence and incidence of late stage
501 age related macular degeneration in the UK. *British Journal of Ophthalmology*.
- 502 9. Mitchell, J., & Bradley, C. (2004). Design of an individualised measure of the
503 impact of macular disease on quality of life (the MacDQoL). *Quality of Life*
504 *Research*, 13(6), 1163-1175.
- 505 10. Mitchell, J., Wolffsohn, J., Woodcock, A., Anderson, S., McMillan, C.,
506 Rubinstein, M., Amoaku, W., & Bradley, C. (2005). + Psychometric evaluation
507 of the MacDQoL individualised measure of the impact of macular
508 degeneration on quality of life. *Health and quality of life outcomes*, 3(1), 25.
- 509 11. Mitchell, J., Wolffsohn, J., Woodcock, A., Anderson, S. J., Ffytche, T.,
510 Rubinstein, M., Amoaku, W., & Bradley, C. (2008). The MacDQoL
511 Individualized Measure of the Impact of Macular Degeneration on Quality of
512 Life: Reliability and Responsiveness. *American Journal of Ophthalmology*,
513 146(3), 447-454.e442.
- 514 12. G., B., Mesbah, M., & Bradley, C. (2011). Metric properties of the MacDQoL,
515 individualized macular-disease-specific quality of life instrument, and newly
516 identified subscales in French, German, Italian, and American populations.
517 *Value in Health*, 1(1524-4733 (Electronic)), 10.
- 518 13. Chakravarthy, U., Harding, S. P., Rogers, C. A., Downes, S. M., Lotery, A. J.,
519 Wordsworth, S., & Reeves, B. C. (2012). Ranibizumab versus Bevacizumab to
520 Treat Neovascular Age-related Macular Degeneration: One-Year Findings
521 from the IVAN Randomized Trial. *Ophthalmology*, 119(7), 1399-1411.
- 522 14. Chakravarthy, U., Harding, S. P., Rogers, C. A., Downes, S. M., Lotery, A. J.,
523 Culliford, L. A., & Reeves, B. C. (2013). Alternative treatments to inhibit VEGF
524 in age-related choroidal neovascularisation: 2-year findings of the IVAN
525 randomised controlled trial. *The Lancet*.
- 526 15. van Reenen, M., & Oppe, M. (2015). EQ-5D-3L User Guide: Basic information
527 on how to use the EQ-5D-3L instrument Version 5.1. Rotterdam.
- 528 16. NICE. (2013). Guide to the methods of technology appraisal. Manchester.

- 529 17. Dolan, P. (1997). Modeling Valuations for EuroQol Health States. *Medical*
530 *Care*, 35(11), 12.
- 531 18. Petrou, S., Rivero-Arias, O., Dakin, H., Longworth, L., Oppe, M., Froud, R., &
532 Gray, A. (2015). Preferred reporting items for studies mapping onto
533 preference-based outcome measures: the MAPS statement. *Quality of Life*
534 *Research*, 1-7.
- 535 19. Wooldridge, J. M. (2009). *Introductory Econometrics: A Modern Approach: A*
536 *Modern Approach*: South Western, Cengage Learning.
- 537 20. Pullenayegum, E. M., Tarride, J.-E., Xie, F., Goeree, R., Gerstein, H. C., &
538 O'Reilly, D. (2010). Analysis of Health Utility Data When Some Subjects Attain
539 the Upper Bound of 1: Are Tobit and CLAD Models Appropriate? *Value in*
540 *Health*, 13(4), 487-494.
- 541 21. Powell, J. L. (1984). Least absolute deviations estimation for the censored
542 regression model. *Journal of Econometrics*, 25(3), 303-325.
- 543 22. Jolliffe D, Krushelnysky B, & Semykina, A. (2000). Censored least absolute
544 deviations estimator: CLAD. *Stata Technical Bulletin*, STB-58(November).
- 545 23. Longworth, L., & Rowen, D. (2011). NICE DSU Technical Support Document
546 10: The Use of Mapping Methods to Estimate Health State Utility Values.
547 Decision Support Unit, SchHARR, University of Sheffield.
- 548 24. Browne, C., Brazier, J., Carlton, J., Alavi, Y., & Jofre-Bonet, M. (2012).
549 Estimating quality-adjusted life years from patient-reported visual functioning.
550 *Eye*, 26(10), 1295-1301.
- 551 25. Sengupta, N., Nichol, M. B., Wu, J., & Globe, D. (2004). Mapping the SF-12 to
552 the HUI3 and VAS in a Managed Care Population. *Medical Care*, 42(9), 927-
553 937.
- 554 26. Brazier, J., Yang, Y., Tsuchiya, A., & Rowen, D. (2010). A review of studies
555 mapping (or cross walking) non-preference based measures of health to
556 generic preference-based measures. *The European Journal of Health*
557 *Economics*, 11(2), 215-225.
- 558 27. Payakachat, N., Summers, K., Pleil, A., Murawski, M., Thomas, J., Jennings,
559 K., & Anderson, J. (2009). Predicting EQ-5D utility scores from the 25-item
560 National Eye Institute Vision Function Questionnaire (NEI-VFQ 25) in patients
561 with age-related macular degeneration. *Quality of Life Research*, 18(9210257,
562 bqm), 801-813.
- 563 28. Dakin, H., Gray, A., & Murray, D. (2013). Mapping analyses to estimate EQ-5D
564 utilities and responses based on Oxford Knee Score. *Quality of Life Research*,
565 22(3), 683-694.
- 566 29. Brazier, J., Connell, J., Papaioannou, D., Mukuria, C., Mulhern, B., Peasgood,
567 T., Lloyd Jones, M., Paisley, S., O'Cathain, A., Barkham, M., Knapp, M.,
568 Byford, S., Gilbody, S., & Parry, G. (2014). A systematic review, psychometric
569 analysis and qualitative assessment of generic preference-based measures of
570 health in mental health populations and the estimation of mapping functions
571 from widely used specific measures. *Health Technol Assess*, 18(34).
- 572 30. Petrou, S., Rivero-Arias, O., Dakin, H., Longworth, L., Oppe, M., Froud, R., &
573 Gray, A. (2015). Preferred Reporting Items for Studies Mapping onto
574 Preference-Based Outcome Measures: The MAPS Statement.
575 *PharmacoEconomics*, 1-7.
- 576 31. Ara, R., & Brazier, J. E. (2010). Populating an Economic Model with Health
577 State Utility Values: Moving toward Better Practice. *Value in Health*, 13(5),
578 509-518.
- 579 32. IVAN Study Team. (2011). IVAN Study Protocol V7.0: NIHR Technology
580 Assessment Programme.

- 581 33. Gray, A. M., Rivero-Arias, O., & Clarke, P. M. (2006). Estimating the
582 Association between SF-12 Responses and EQ-5D Utility Values by
583 Response Mapping. *Medical Decision Making*, 26(1), 18-29.
584